

МИТРЕНИНА ОЛЬГА ВЛАДИМИРОВНА

к.ф.н., доцент кафедры математической лингвистики, Санкт-Петербургский государственный университет, o.mitrenina@spbu.ru

Автоматизация нулевого и первого этапа разметки рук при создании мультимодального корпуса разговорной речи

Ключевые слова: мультимодальный корпус; разговорная речь; нейросетевая разметка жестов; корпусная лингвистика

В докладе описывается серия экспериментов по нейросетевой разметке жестикуляции для мультимодального корпуса разговорной речи. Значимость проблемы обусловлена следующим. В живом общении важны не только слова, но мимика, жесты, интонация, движения глаз, проксемика и еще какие-то параметры, о которых мы, возможно, даже не знаем. Если изучение интонации высказываний давно уже стало частью лингвистических исследований, то изучение жестикуляции становится доступным для систематического изучения лингвистам только сейчас, хотя никогда не отрицались ни значимость жестикуляции и мимики для речевой коммуникации, ни различия в соответствующих привычках между носителями разных языков. Еще недавно исследовать эти параметры (и их взаимосвязь) с помощью корпусных методов возможности не было по причине слабости доступных тогда компьютеров. Однако сейчас производительность компьютеров существенно увеличилась, и поэтому большую популярность приобретают мультимодальные корпуса, которые содержат видеозаписи коммуникаций с разметкой различных каналов: текст, интонация, жесты, мимика, движения глаз и пр. Это открывает совершенно новые перспективы для исследования невербальных аспектов речевой коммуникации, и нам только предстоит осознать, насколько эти перспективы велики.

В настоящей работе эксперименты проводились с записями корпусом «Рассказы и разговоры о грушах», который создается в Институте языкознания РАН. Корпус содержит записи естественной коммуникации между несколькими собеседниками, а также их вокальную, референциальную, кинетическую и окулomotorную аннотацию [Кибрик, Федорова 2018]. Корпус включает 24 записи общей длительностью около 9 часов, но в открытый доступ выложены три записи 2015 года, каждая из которых включает 18 медиафайлов, каждый из которых снабжен многоуровневой аннотацией. Кинетическая аннотация корпуса, выполненная в программе ELAN, содержит аннотацию жестов рук и жестов головы.

В ходе представленного в докладе эксперимента видеозаписи корпуса были проанализированы и размечены с помощью искусственных нейронных сетей. Тестировались наиболее популярные среди разработчиков в настоящее время модели глубокого обучения MediaPipe, MoveNet и YOLOv8-pose. Лучший результат при распознавании пальцев на видео показало использование библиотек [MediaPipe Solutions].

Анализируемое видео было разбито на фреймы (каждой секунде видео соответствовали 30 фреймов). Для каждого фрейма фиксировалась временная метка (timestamp) и с помощью нейронной сети определялись координаты ключевых точек обеих рук говорящего: запястья и кончиков трех пальцев — мизинца (PINKY), указательного пальца (INDEX) и большого пальца (THUMB). Данные были собраны в форматах CSV и Excel. После этого на основании анализа расстояний между ключевыми точками проводился поиск аномальных значений в координатах. Для этого было определено пороговое значение для допустимого расстояния между запястьем и кончиками каждого из трех пальцев. Таким образом, были выделены жесты, представляющие трудность для автоматического анализа.

Следующим этапом для каждой временно метки диалога были вычислены скорость и направление движения каждой ключевой точки обеих рук. Для этого для каждого момента времени учитывались x- и y-координаты этих точек в предыдущий языковой момент

Полученный результат позволяет автоматизировать нулевой этап разметки [Литвиненко и др. 2018], который раньше составлялся вручную. Этот этап включает ответы на следующие вопросы:

1. Много ли у говорящего жестов?
2. Много ли у говорящего крупных (структурированных) адапторов?
3. Много ли у говорящего мелких (неструктурированных) адапторов?
4. Какова типичная продолжительность покоя между цепочками движений?
5. Какова типичная скорость движений?
6. Какова типичная амплитуда движений?
7. Каковы типичные нейтральные положения?
8. Есть ли частотные рекуррентные (повторяющиеся) жесты?
9. Есть ли частотные повторяющиеся адапторы?

Типичная амплитуда и типичная скорость движения определяются без использования нейронных сетей. Для остальных показателей требуется небольшой обучающий корпус, содержащий элементы разметки мануальных жестов и адапторов — не являющихся коммуникативно значимыми движений, направленных на поддержание физического комфорта говорящего. Для этого используются фрагменты созданной вручную аннотации движений рук нескольких информантов. Дальнейшее развитие предложенного подхода позволяет перейти к нейросетевой разметке типов жестов и адапторов.

Полученная разметка выравнивается с текстовой расшифровкой диалогов и сегментной аннотацией, выполненной в программе Praat, а также с транскриптами с указанием временной динамики вербальных элементов [Коротаев 2019]. Даже на этом этапе подобное выравнивание позволяет выявить определенные корреляции между жестами, интонацией и различными лексическими, грамматическими и другими аспектами устной речи.

На примере моделей, использованных в данном эксперименте, можно убедиться, что исследование невербальных аспектов речевой коммуникации может и должно в настоящее время выйти на новый уровень, который обеспечит гораздо более «стереоскопичное», если можно так выразиться, понимание разговорной речи, не ограниченное одним лишь вербальным ее аспектом.

Литература:

1. Кибрик А. А., Федорова О. В. 2018. An empirical study of multichannel communication: Russian Pear Chats and Stories // Психология. Журнал Высшей Школы экономики №15 (2). С. 191–200.
2. Коротаев Н.А. «Рассказы и разговоры о грушах»: принципы вокальной аннотации. Версия 10.01.2019. <http://multidiscourse.ru>
3. Литвиненко А.О., Николаева Ю.В., Кибрик А.А., Федорова О.В. «Рассказы и разговоры о грушах»: аннотирование движений рук. Версия 14.12.2018. <http://multidiscourse.ru>
4. MediaPipe Solutions. URL: <https://github.com/google-ai-edge/mediapipe>

Исследование подготовлено при поддержке СПбГУ, шифр проекта 124032900006-1

...руки срывают большую зеленую грушу...